Adapting Animal Models to Assess Sufficiency of Fluid Resuscitation in Humans (Student Abstract)

Ryan Schuerkamp[1,2*], Xinyu Li[1], Brian Kunzer[1], Leonard S. Weiss[3], Hernando Gómez[3], Francis X. Guyette[3], Michael R. Pinsky[3], Artur Dubrawski[1]

[1]Auton Lab, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[2]College of Engineering and Computing, Miami University, Oxford, OH 45056, USA
[3]School of Medicine, University of Pittsburgh, Pittsburgh, PA 15213, USA
{rschuerk, xinyul2, bkunzer, awd}@andrew.cmu.edu, {weissls2, gomezh, guyefx}@upmc.edu, pinsky@pitt.edu

Abstract

Fluid resuscitation is an initial treatment frequently employed to treat shock, restore lost blood, protect tissues from injury, and prevent organ dysfunction in critically ill patients. However, it is not without risk (e.g., overly aggressive resuscitation may cause organ damage and even death). We leverage machine learning models trained to assess sufficiency of resuscitation in laboratory animals subjected to induced hemorrhage and transfer them to use with human trauma patients. Our key takeaway is that animal experiments and models can inform human healthcare, especially when human data is limited or when collecting relevant human data via potentially harmful protocols is unfeasible.

Introduction

Severe blood loss is one of the primary causes of possibly preventable death in both civilian and military trauma (Alam and Rhee 2007). Fluid resuscitation is an initial treatment frequently employed to treat shock, restore lost blood, and regain organ function in critically ill patients; it is the most common intervention in acute medicine (Myburgh 2015). However, its application is not without risk. An overly aggressive resuscitation may result in fluid overload, organ damage, and even death (Myburgh 2015). Thus, clinicians, as well as rescuers in the field, must conduct it carefully.

Previous work developed a machine learning-based approach for fluid resuscitation leveraging data collected from pigs during laboratory experiments and only non-invasive features (Li, Pinsky, and Dubrawski 2022). Specifically, it predicted sufficiency states at each resuscitation time step, where sufficiency refers to the cardiovascular parameters necessary for adequate blood flow to meet the metabolic demands of the body without overtly failing (i.e., a subject does not need more aggressive resuscitation during a sufficient state). However, real-world application requires validation in humans, a challenging task due to the highly controlled laboratory setting for the pigs and human interpatient diversity. We extend prior work in two significant ways: (1) we apply the optimized aggregation of predictions pig model from Li, Pinsky, and Dubrawski (2022) and develop additional models to predict resuscitation decisions in two datasets of human patients airlifted to a hospital experiencing severe blood loss and (2) we apply a domain adaptation approach to augment the pig model with human data, improving performance on one of the human patient datasets and demonstrating that pig data may inform human models.

*Phone Number: +1 952-250-2249

Methods

Human Data

We use two datasets containing both invasive and non-invasive vital sign data of trauma patients airlifted to a hospital: one of 62 patients given an arterial line and the second set of 98 patients without an arterial line. We employ a similar window and feature extraction approach as in Li, Pinsky, and Dubrawski (2022). We extract resuscitation labels using the timestamps of blood, fluid, and vasopressor administration. Specifically, we define five minutes before and after norepinephrine or epinephrine (two common vasopressors) administration and 15 minutes before and after fluid or blood administration as the ground truth labels for resuscitation periods, and consecutive blood or fluid administrations are combined into one if given within 20 minutes of each other. After extracting ground truth labels, we had 9639 periods where the patient was not resuscitated and 1034 resuscitation periods for the 62 patients given an arterial line and 30 features, a subset of the 42 initially used for the pig models. For the 98 patients without an arterial line, we had 10785 no resuscitation and 1241 resuscitation periods with 27 features. We eliminated the three mean arterial pressure (MAP) features because they were not available for the second set of patients as they did not have arterial lines.

Model Fitting

We apply four random forest models to the human data and evaluate them using leave-one-subject-out cross-validation. First, we fit one using only the human data. Second, we correct the class imbalance using SMOTE (Chawla et al. 2002) (an oversampling method) and then fit the model. Third, we apply transfer learning to the pig model to tailor it to the human data using the MIX algorithm from Segev et al. (2016). Finally, we apply SMOTE and refit the previous model.

Converting Sufficiency Predictions into Resuscitation Decisions

The three parameters used to convert sufficiency predictions into resuscitation decisions on human data are first calibrated leveraging the pig model, a dataset of 30 pigs sedated for induced hemorrhagic shock and fluid resuscitation, and leave-one-subject-out cross-validation. First, we calibrate the decision threshold for sufficiency (i.e., the model prediction score required to consider a prediction sufficient) and allow it to vary from zero to one in one percent increments. Then, we calibrate the number of consecutive sufficiencies and insufficiencies required to stop and start resuscitation, respectively, and allow them to vary from zero to nine. For example, if a decision threshold of 0.6 and zero consecutive sufficiencies and insufficiencies are selected, any prediction with a probability of sufficiency greater than or equal to 0.6 will end the ongoing resuscitation, and any prediction with a probability of sufficiency less than 0.6 will (re-)start resuscitation.

Evaluation

Accuracy is not an optimal metric due to the severe class imbalance, so we evaluate models using the overlap between the ground-truth and the predicted resuscitation periods:

$$overlap = \frac{\frac{overlap\ sum}{annotated\ sum} + \frac{overlap\ sum}{vote\ sum}}{2}, \tag{1}$$

where $annotated\ sum$ is the total time labeled as sufficient/do not resuscitate or insufficient/resuscitate by the ground truth, $vote\ sum$ is the total time predicted by the model, and $overlapsum$ is the total time these two sets overlap. $\frac{overlap\ sum}{annotated\ sum}$ assesses how much ground truth was successfully predicted, and $\frac{overlap\ sum}{vote\ sum}$ evaluates how much of the model predictions overlap with the ground truth. Thus, it balances identifying the ground truth with making precise predictions and assumes a value between zero and one, with a value of one indicating the predictions perfectly overlap with the ground truth and zero denoting no agreement.

Results

The optimal parameters to convert sufficiency predictions into resuscitation decisions are a decision threshold of 0.77 and zero consecutive insufficiencies and sufficiencies. For the humans with an arterial line, the transfer learning model without SMOTE has the highest resuscitation overlap of 0.368 ± 0.033 (Table 1), so it is the best at identifying resuscitation periods. The transfer learning model with SMOTE has the highest no resuscitation overlap of 0.63 ± 0.027. For the humans without an arterial line, the human only model without SMOTE has the highest resuscitation overlap of 0.593 ± 0.049. The human only model with SMOTE has the highest no resuscitation overlap of 0.587 ± 0.020.

	Arterial Line		Without Arterial Line	
Model	No Res Overlap	Res Overlap	No Res Overlap	Res Overlap
Human Only	0.023 ± 0.018	0.002 ± 0.001	0.089 ± 0.027	**0.593 ± 0.049**
Human Only SMOTE	0.598 ± 0.023	0.272 ± 0.028	**0.587 ± 0.02**	0.214 ± 0.031
Transfer Learning	0.493 ± 0.014	**0.368 ± 0.033**	0.556 ± 0.018	0.186 ± 0.028
Transfer Learning SMOTE	**0.63 ± 0.027**	0.243 ± 0.028	0.564 ± 0.019	0.182 ± 0.028

Table 1: Mean and standard error of the results for the human patients based on leave-one-subject-out cross-validation. Note, res stands for resuscitation.

Discussion

Fluid resuscitation is widely used to treat the adverse effects of severe bleeding and the subsequent shock; however, it is challenging. Inadequate resuscitation may fail to treat shock if present or prevent advancement to shock for patients at risk, and over-resuscitation may damage organs or even cause death. We built on previous work to successfully convert sufficiency predictions into resuscitation decisions, demonstrating the potential of our method to be adapted into an automated control system for fluid resuscitation administration. We applied a model trained on pig data via transfer learning to two human datasets, finding the transfer learning approach had the highest performance at identifying both resuscitation and no resuscitation desired states for the human patients with an arterial line and MAP data available but not the patients without an arterial line and MAP data. The presence of the three MAP features in the pig dataset and patients with an arterial line and their absence in the patients without an arterial line likely contributed to this discrepancy because low blood pressure is a common trigger for fluid resuscitation. Our results demonstrate that models developed from data collected in controlled laboratory animal experiments may inform complex clinical tasks involving human patients using a machine learning-driven approach.

Acknowledgements

This work was partially supported by the Research Experiences for Undergraduates program under NSF grant 1730147, by the University of Pittsburgh Center for Research Computing, RRID:SCR_022735 using the HTC cluster supported by NIH award S10OD028483, by NIH award R01HL141916, and by the U.S. Army awards W81XWH-19-C0083 and W81XWH-19-C0101.

References

Alam, H. B.; and Rhee, P. 2007. New developments in fluid resuscitation. *Surgical Clinics*, 87(1): 55–72.

Chawla, N. V.; Bowyer, K. W.; Hall, L. O.; and Kegelmeyer, W. P. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16: 321–357.

Li, X.; Pinsky, M. R.; and Dubrawski, A. 2022. Automated Assessment of Cardiovascular Sufficiency Using Non-Invasive Physiological Data. *Sensors*, 22(3): 1024.

Myburgh, J. A. 2015. Fluid resuscitation in acute medicine: what is the current situation? *Journal of Internal Medicine*, 277(1): 58–68.

Segev, N.; Harel, M.; Mannor, S.; Crammer, K.; and El-Yaniv, R. 2016. Learn on source, refine on target: A model transfer learning framework with random forests. *IEEE transactions on pattern analysis and machine intelligence*, 39(9): 1811–1824.